MARINE RECREATIONAL INFORMATION PROGRAM

MRIP Data Management Standard (MDMS) Database

PROJECT REPORT

DMSWG-1 Identify and consolidate information on existing recreational datasets

Project leader: Vivian Matter

Project members: Gregg Bray, Tina Chang, Lauren Dolinger Few, Chad Hanson, Gretchen Jennings, Bruce Joule, Kathy Knowlton, Anjel Lewis, Dennis O'Hern, Risa Oram, Michael Quach, Scott Sauri, Tom Si, Wade Van Buskirk, and Patty Zielinski

Data Management and Standards Workgroup
Chair: Kathy Knowlton
Vice-Chair: Gregg Bray

November 4, 2008

Table of Contents

1 Overview of DMSWG-1

The Data Management and Standards Workgroup (DMSWG) is charged, in part, with developing data collection standards, protocols, and data access interfaces for the Marine Recreational Information Program (MRIP). DMSWG project 1 (DMSWG-1) takes the first step in accomplishing this task by identifying and summarizing the existing fisheries-dependent recreational, for-hire, and highly migratory species (HMS) programs. After summarizing this information the DMSWG will be able to categorize a set of common data elements that currently exist in the catch and effort data. In addition, the data collected will be a reference for many of the other tasks the DMSWG will need to complete (establishing protocols, standardizing data collection, etc).

2 MRIP Data Management Standard (MDMS)

2.1 Purpose

When the MRIP Data Management and Standards Work Group (DMSWG) Project 1 (DMSWG-1) plan was initially put together, the DMSWG intended to record all the program metadata in a Microsoft Word (Word) template. The project leader would then review all the Word documents and compile a final report by copying and pasting information into one master Word document.

However, once the DMSWG began to work on the template, it became apparent that storing the metadata in Word documents was not an efficient solution. After evaluating several options, the DMSWG decided that it was necessary to build a web-based, database driven application to store the data. This solution provided the following benefits:

a. A standard interface through which all project participants could enter metadata regardless of location (assuming Internet connectivity).
b. A centralized solution that would make it easy to tweak the interface as it was used without having to send out new versions of the template to project participants, requiring them to port previous data into the latest version.
c. A well designed database that would allow far more robust analysis and reporting options than a Word document could provide.
d. Mechanisms for quality control and process workflow that would not be possible using Word documents.
e. A mechanism to allow multiple people to access the metadata at the same time, which allows for a much higher level of collaboration (i.e. the project leader would not have primary responsibility, while the other project participants wait for a massive Word report to be compiled).
f. A base system that could be built upon as program metadata were further developed.

It should be noted that before beginning development on the MDMS system, the DMSWG considered utilizing InPort for initial data entry. InPort is the national Fisheries Information System (FIS) metadata tool. However, it was determined that given the need for very specific data fields and the time constraints associated with DMSWG Projects 1 and 2, it was necessary to build a bridge system to accommodate the initial input and analysis of metadata by DMSWG project participants with the understanding that these metadata would ultimately be input into InPort for wider use.

2.2 Scope

The MDMS includes information about current fisheries-dependent recreational, for-hire, and highly migratory species (HMS) programs nationwide. Programs from the continental United States, Alaska, Hawaii and the Pacific Island and Caribbean territories are included. The types of programs covered include surveys, logbooks, catch card, tournament, tagging programs and others.

2.3 Functionality

The MRIP Data Management and Standards (MDMS) system is a metadata collection interface. The MDMS system was built specifically for the Marine Recreational Information Program (MRIP) Data Management and Standards Work Group (DMSWG) Project 1 (DMSWG-1), which entails identifying and consolidating information on existing recreational datasets. Data entered into the MDMS database will be used by work group members to analyze existing data collection programs and datasets to develop recommendations for minimum data elements and standards in the new MRIP system.

The interface contains the following features:

- Login page that requires a username and password for access to the application
- User access rights at the module level, record level and field level based on user role
- A robust yet easy to use data entry and review form containing sixty fields divided by information category into eight primary tabs and three secondary tabs
- Search functionality (to locate records)
- A glossary of fisheries terminology
- A help screen
- Account/user management functionality (including role assignment)
- Program assignment functionality (to associate users with programs)
- Program inspection functionality (to determine record completeness)
- Program review functionality (for verification of data)
- Export to .csv functionality (for ad hoc report generation)
- Change history functionality (records all changes to programs)

3 Data Collected

An important goal for this project was to collect data about the current fishery-dependent data collection programs in a uniform manner. The same type of information was collected about each separate program. This will facilitate program comparisons and aid in the development of standards and protocols for the new MRIP system. As previously mentioned, sixty fields of data were collected and divided by information category into eight primary tabs and three secondary tabs outlined below. A complete list of data collected can be found in Appendix A.

- Program- general information about each program such as name, type, program description, and design statement
- Contact(s)- one or more contacts for the program
- Coverage(s)- program's temporal and geographical coverage [The first part describes the range of years and geographical areas that are covered by the program such as, "Gulf Coast states from 1980 to the present." The second part lists exceptions to these ranges such as, "No coverage for Alabama in 1990"]
- Info Manage- data auditing procedures, QA/QC processes, availability of program documentation, etc.
- Data Collection- interviewer training protocols, the type of data collected, who collects it, when it is collected, and how it is collected
- Data Processing- data entry method, the processing frequency, and data transformation processes
- Data Access- types of user interfaces, any confidentiality issues, timeliness of data output, and descriptions of user access levels
- Data Elements- set up in a tree structure and stores the name, description, and other data about each data set in the program, each data entity in each data set, and each data attribute in each data entity
    - Data Sets- e.g. Access-Point Intercept Data
        - Data Entities- e.g. i3 files (available catch)
            - Data Attributes- e.g. sp_code (species of fish)

4 Data Programs

Following is a list of programs by region that have been inputted into the MDMS database as of the date of this report. Each program was approached with the intent to fully document the program in a manner consistent across all programs. However, due in part to varying levels in the availability and cooperation of program representatives and also in part to time and resource constraints, some programs are more fully documented than others.

4.1 Alaska

Alaska Saltwater Logbook Program
Alaska Statewide Harvest Survey
Southcentral Alaska Halibut and Groundfish Harvest Assessment Project
Southeast Alaska Marine Harvest Studies

4.2 Pacific Islands

Access-point intercept survey (Hawaii)
American Samoa Boat-based Creel Survey
American Samoa Shore-based Creel Survey
CNMI Boat-based Creel Survey
CNMI Shore-based Creel Survey
Guam Boat-based Creel Survey
Guam Shore-based Creel Survey
Hawaii Fishermen Reporting System
Hawaii Recreational Meta-Data Project (Tournaments and Clubs)
Hawaii Ulua and Papio Tagging Project
Maunalua Bay Pakini Survey (Hawaii)

4.3 Pacific Coast

Access-point intercept survey (Pacific coast)
BOAT-PC On-board survey of Party and Charter boats for catch location and depth
CA-ALS California Angler license survey for effort
CA-LOGS California logbook program of Passenger fishing boats for effort and catch.
CRFS-BB California Recreational Fishers Survey of Beaches and Banks for catch
CRFS-MM California Recreational Fisheries Survey of Man Made Structures
CRFS-PC California Recreational Fisheries Survey of Party and Charter boats for catch
CRFS-PR1 California Recreational Fisheries Survey of Primary Launch Ramps
CRFS-PR2 California Recreational Fisheries Survey of Secondary Launch Ramps
DISCARD Intercept sample for discard fish size and discard disposition
OR-ALS Oregon Angler License Survey for effort
ORBS Oregon Boat Survey for effort and catch
OSP-PC California Ocean Salmon Project survey of Party and Charter Boats
PCPS California Party and Charter boat Phone Survey
PSBS Washington Puget Sound Boat Survey for catch
Pacific RecFIN Estimates
SEB Oregon Shore and Estuary Boat Survey for catch
WA-ALS Washington Angler License Survey for effort
WA-OSP Washington Ocean Sampling Program for effort and catch

4.4 North and Mid-Atlantic

Connecticut Volunteer Logbook Program
Delaware Fishing Effort on Delaware Reef Sites

Delaware Recreational Clam Survey
Delaware Recreational Gill Net Survey
Highly Migratory Species Catch Card Program – Maryland
Improving the Accuracy of Length and Weight Information for Angling Category Atlantic Bluefin Tuna
Large Pelagics Intercept Survey
Large Pelagics Telephone Survey
Maine Volunteer Logbook Program
Maryland Coastal Bays Angler Survey
Maryland General Tourist Survey
Maryland Offshore Recreational Angling Survey
Maryland Spring Striped Bass Creel Survey
Massachusetts Sportfishing Tournament Monitoring Program
NMFS Northeast Region Vessel Trip Reporting (VTR) / Logbook program
New Hampshire Smelt Creel Survey
New Hampshire Striped Bass Volunteer Angler survey
New Jersey Recreational Blue Crab Survey
New Jersey Striped Bass Volunteer Angler Survey
New York Striped Bass Cooperative Angler Program
Virginia Web Based Survey of Recreational Fishing Activity

4.5 South Atlantic and Gulf of Mexico

Access-point intercept survey (Gulf of Mexico)
Everglades National Park Creel Survey
Everglades National Park Professional Guide Logbook
Florida - Bay Scallop Survey
Florida - Spiny Lobster Recreational Survey
Georgia Carcass Recovery Project
GulfFIN Biological Data Collection Program
Highly Migratory Species Catch Card Program - North Carolina
North Carolina Cooperative Striped Bass Creel Survey
Southeast Headboat Survey and Logbook Program (SEHB)
Texas Parks and Wildlife Private Boat Angler Survey

4.6 Caribbean

There are currently no programs for the Caribbean entered into the MDMS system.

4.7 Multi-Regional programs

Access-point intercept survey (Atlantic coast and Caribbean)
Atlantic HMS Automated Landings Reporting System (ALRS)
Catch, Effort and Participation Estimates
Coastal Household Telephone Survey (CHTS)
For-Hire Survey (Atlantic and Gulf)
International Billfish Angler Survey
Recreational Billfish Survey

Appendix A: MDMS Help Text

Location: MRIP Collaboration Tool
Content: MDMS Entity Data Glossary

GETTING STARTED:

http://www.st.nmfs.noaa.gov/mdms/html/mdms.html

First you will get the login window for the MDMS

```
Login = firstname.lastname
password = [supplied by Scott]
```

You are encouraged to change your password.

Once in the MDMS:

Available functionality and program record access depends on your user role:

Guest: access to all functionality except "Manage Accounts" and "Inspect Programs" with read-only access to assigned programs.
Reviewer: same as guest, with edit access to "Reviewer" section on Program tab
Manager: access to all functionality except "Manage Accounts" with full access to all programs
Admin: access to all functionality, with full access to all programs

The top is for choosing which mode you want to use.
The left and right windows will change depending on which mode you select.
The vertical separator is dynamic, you can move it to provide more/less room as needed.
The GLOSSARY will open another window with NOAA's standard glossary.

In 'Manage Programs' mode:

The left side is for searching and listing of programs.
The right side is for entering and editing program metadata.

LEFT SIDE:
The search window is not case-sensitive, but only searches the "Program Name" field, not the associated metadata.

In the list of programs, clicking the edit icon ( ) selects the program and populates the editing tabs on the **RIGHT** side. If you do not see the programs listed, click the "manage programs" link to refresh screen. If you want to clear a search, 'reset' the criteria, and 'search' again.

Everything else pertains to the contents of the window on the RIGHT SIDE of the page.

RIGHT SIDE:

Suggestion: Some of the fields involve entering large amounts of text. You may want to type this information FIRST into a Word or other text document, then copy and paste into MDMS. If odd characters (e.g. ¿) appear in your text, don't worry about them, we can correct that on the database end of things.

PROGRAM TAB:

Program Name: A name unique to your program.

Program [Sub] Type: If you select a "Type" (i.e. survey) that has a subtype, that option will appear. If you select a type w/o a subtype, that option will disappear.

Program Record Status: This pertains to the status of entering the program metadata into MDMS. Until you are completely done, leave as 'Incomplete'

Responsible organization: Who is responsible for the data (owner)? Hold CTRL to select multiple orgs.

Funding Source: Who provides the funds for the program? Hold CTRL to select multiple orgs.

Program/Project Website: URL link (e.g. www.nmfs.noaa.gov)

Program Description: General information about the program

Design Statement: General statement on what data the program is designed to provide (e.g., estimates by Wave, Dual frame, sampling strategy, etc.)

Record Notes: A catch-all, primarily to assist coordinator/proxy when inputting metadata, such as what tabs they still needed to populate, etc. "I stopped entering data on the Coverages Tab" or "still need to add more to Program Description"

Reviewer Area

The "Reviewer Notes" field and "Save Reviewer Notes" button: Used by the reviewer to document any changes that need to be made to the program record (e.g. corrections, supplemental information).

"Save and Verify" button: Used by the reviewer to save the reviewer notes and change the program record status to "Verified".

CONTACT(S) TAB:

The top of the page is for adding and editing contacts, the bottom will list the contacts after they have been saved.

Role: Describes how a person is involved with program, so if you need to contact someone for data, talk to "A", or if you have questions about survey protocols, contact "B" (e.g. data manager, program manager, division chief, COTR)

COVERAGE(S) TAB:

Program Coverage area

This section describes the overall program in temporal and geographic terms. If a program has been running for 10 years, but has a couple of months that got skipped in the middle somewhere, you can describe the skipped months (i.e. "exception") in the lower section of the page

Program Start Date and End Date: If the program is on-going, just leave the end date blank.

Program Coverage Notes: At a minimum, list out each individual state (FL or Florida). You can ALSO include subregions and regions (e.g. North Atlantic). This field is for searching purposes so DO NOT list a range (e.g. Maine to Florida) as it makes it almost impossible to tell (using standard search queries) that North Carolina is included. If a state was ever included in the program, list it. If it was only included 1 out of 10 years, include it… you will list the other 9 years in the EXCEPTIONS section below.

Exceptions area

This area is where you identify all the exceptions to the Program Coverages described above. The EXCEPTION NOTES description should be clear/unique enough to separate from other exceptions.

INFO MANAGE TAB:

Available Documentation: List available documentation, like a "User's Manual" or "Procedures Manual". This is just a list of the documents, not the content. You can add a URL, but that does not mean you are exempt from filling out information elsewhere in the MDMS.

Data Auditing/QA/QC processes: Describe the processes in place.

Information modification policies: Are they any policies in place for editing historical data, if so… describe.

DATA COLLECTION TAB:

What data are collected: Angler, biological, and/or boat data? Catch, Effort…

Who collects the data: Responsible party out in the field, not the overall agency charged with collecting/archiving/processing the data.

Data Collection Frequency: Text description, such as "The week before and after the end of the Wave (every two months)" or "Monday-Friday every week of the year".

Methodologies/Protocols: Paste/type in text from manuals, etc. Include variations over time/geography.

Interviewer Training: What types of QA.QC protocols are in place for data collection, how are interviewers trained, how are interviewers evaluated?

Data Collection Method: Use CTRL to select more than one

Any Coding systems used?: Use CTRL to select more than one

DATA PROCESSING TAB:

Data Entry Method: How does the data get from the [paper] "form" to the database. If someone sits at a computer and types it in = Manual. OCR="Optical Character Recognition".

Data Processing Frequency: How often to you go through the above procedures? Every month, once a year, on a wave (two-month) basis?

General Data Transformation Processes: If the "native" format of the data is EXCEL, but you convert it to SAS.. describe how.

Data Processing Details: What do you do with the data once you have it? Run SAS code, produce Excel tables, etc. Are there any expansion techniques applied? Do you just provide the data to someone else?

DATA ACCESS TAB:

Confidentiality: What controls/limits are in place?

Description of access level, user roles: Who is allowed access to the data (e.g. contractor, NMFS)

User interface(s): How and where can someone gain access to the data/products of this program?

Timeliness: How often do you release the data/products? (e.g. Preliminary data are available 45 days after the end of the wave, final data are generally available by April 15 of the following year.")

DATA ELEMENTS TAB:

Glossary of terms:

DATA SET=name of a group/collection of files/database
DATA ENTITY=table/file name
DATA ATTRIBUTE=field/column/variable

DATA SETS [database] TAB:

Data Set Name: Database name or schema, or name for collection of related file (e.g. Intercept files)

Data Set Description: General description of the data contained in file. Specific details (filenames, variables and such) will be documented in the DATA ENTITIES/ATTRIBUTES tab.

Data Set Format: How are these data stored? If "Other" is selected a text box will appear to type in more detail. ***FYI – SAS files are flat files.***

Data Set Application: What application/software is used to store the data. If "Other" is selected a text box will appear to type in more detail.

Application Version: e.g. 10g, XP, 9.1.3

Data Set Current As Of: most current date.

Once a dataset has been created, a list will appear at the bottom. After a data set has been created or selected, the DATA ENTITIES tab will appear.

DATA ENTITIES [tables] TAB:

Data Entity Name: Name of the table/file

Data Entity Type:

Data Entity Description: describe the data generally ("Angler trip data"). Specific details of the variables and codes are described in the DATA ATTRIBUTES.

Once a data entity has been created, a list will appear at the bottom. After a data entity has been created or selected, the DATA ATTRIBUTES tab will appear.

DATA ATTRIBUTES [variables] TAB:

In order to add attributes about an entity, the entity must be selected from the list (fields should populate the DATA ENTITIES tab before moving on to DATA ATTRIBUTES tab).

Data Attribute Name: Variable/field name

Data Attribute Type: format of attribute

Length: May refer to number of bytes used for storing the attribute (as with SAS) or number of characters in the attribute (Oracle).

Precision: For numeric attributes, the precision (number of digits) with which a value is expressed.

Data Attribute Required: Optional attribute/null allowed?

Part of Primary Key: Is the attribute a unique identifier, or combined with other attributes, does it create a unique identifier for the record?

Part of a Foreign Key: Is this attribute a primary key in another table?

Data Attribute Description: Describe the attribute, include label or further details.

Related Tables or Code: List of lookup table(s), or list of codes used as values for this attribute.

Once an attribute has been created, a list will appear at the bottom. The "visible columns" link allows you the edit the columns displayed in the list. Clicking the edit icon selects the attribute and populates the editing fields above.

In 'Export' mode

This module allows for the export of MDMS data to an .xls file for the purpose of ad hoc reporting. The left side are search criteria and report results, the right will populate with data upon program selection.

The first component is a search with options to search by program name or program assignee (note: the program assignee search is a string search on the username, so if you are looking for all records assigned to me (Scott Sauri), typing "scott.sauri" or just "scott" will work, but typing "scott sauri" will not work)

The second component is a list of program record fields to be included in the report. Each field selected will be a column in the report. Contacts and coverages are not included as options.

The third component is a list of search results. Selecting a record by clicking on the “pen and paper” icon to the left of the Program Name will load the record into the right side of the screen. The report is exportable to .xls format.

In ‘Inspect Programs’ mode:

This module is an ad hoc reporting tool that is being used by the team members to review program records for completeness and consistency. It has the same basic components as Export mode.